Supervised Spectral Subspace Clustering for Visual Dictionary Creation in the Context of Image Classification

Imtiaz Masud Ziko, Elisa Fromont, Damien Muselet, Marc Sebban

Supervised Spectral Subspace Clustering for Visual Dictionary Creation in the Context of Image Classification

Imtiaz Masud Ziko, Elisa Fromont, Damien Muselet and Marc Sebban
Laboratoire Hubert Curien, Saint Etienne, France
ziko.iut@aiub.edu, {elisa.fromont, damien.muselet, marc.sebban}@univ-st-etienne.fr

Abstract

When building traditional Bag of Visual Words (BOW) for image classification, the K-means algorithm is usually used on a large set of high dimensional local descriptors to build the visual dictionary. However, it is very likely that, to find a good visual vocabulary, only a sub-part of the descriptor space of each visual word is truly relevant. We propose a novel framework for creating the visual dictionary based on a spectral subspace clustering method instead of the traditional K-means algorithm. A strategy for adding supervised information during the subspace clustering process is formulated to obtain more discriminative visual words. Experimental results on real world image dataset show that the proposed framework for dictionary creation improves the classification accuracy compared to using traditionally built BOW.

1. Introduction

Bag of visual words (BOW) method [7] has marked significant advancements in the image classification applications. The method involves extracting a large set of discriminative local descriptors (such as SIFT, Color-SIFT [1] etc.) from the training images and learning the visual dictionary by applying clustering algorithm, usually the K-means in the descriptor space. Finally, the cluster representatives are called the visual words of the visual dictionary. However, the K-means algorithm considers all the feature dimensions while calculating the distance during estimating clusters. As a consequence, actual nearest neighbor estimation can get affected by having noisy information from irrelevant feature dimensions [4, 13]. Image descriptor domain is a complex structure having high dimensionality where feature descriptors are assumed to lie in a union of linear or nonlinear subspaces. Thus, applying the K-means algorithm often fails to capture the underlying structure of the descriptor space.

Alternatively, we have experimented to extract the visual words according to the relevant subspaces by using suitable subspace clustering methods. Subspace clustering methods were not explored very well for this specific problem related to the BOW model for image classification. There has been a recent work [5] where they worked to get the nonlinear subspaces from the raw image space by using Restricted Boltzmann Machine (RBM) [11]. In the comparison of results, they incorporate spatial information using the SPM method [14] during the classification. In contrast, we aim to have a fair comparison between simple BOW model using K-means and using subspace clustering approach without using additional tweak of adding spatial information in the built dictionary.

Spectral clustering methods [19] have shown a tendency to perform better in comparison to other existing unsupervised methods for object discovery as reported in [17]. In general, the spectral clustering methods construct a *similarity matrix* $W \in \mathbb{R}^{N \times N}$ by estimating pairwise similarity among all the N data points. Let $G(V, E)$ be an undirected graph where V is the set of N vertices and E is the set of weighted edges. Ideally, $W_{ij} = 1$ if points i and j are connected and $W_{ij} = 0$ otherwise. Finally, the K clusters are achieved by applying the K-means to the subset of K eigenvectors of the Laplacian matrix $L \in \mathbb{R}^{N \times N}$ formed from W. There are several ways to get the Laplacian matrix such as: the unnormalized Laplacian $L = diag(W\mathbf{1}) - W$, the normalized Laplacian $L_{sym} = diag(W\mathbf{1})^{-0.5} L diag(W\mathbf{1})^{-0.5}$ and $L_{rw} = diag(W\mathbf{1})^{-1} L$ where $\mathbf{1}$ is the vector of all 1's.

Spectral subspace clustering methods attempt to build the *similarity matrix* W such that only the data points coming from the same subspace will be connected. The rest of the steps remain the same as the spectral clustering method. We compare – SSC [8], SMCE [9], SLBF [20] and SCC [6] for our experiments due to their promising performances reported in the literature for different subspace segmentation problems [18]. Initially we assess these methods on synthetic dataset and based on their performances, pick three of them – SSC, SMCE and SLBF for our proposed framework. The main contribution of this paper is that we propose a

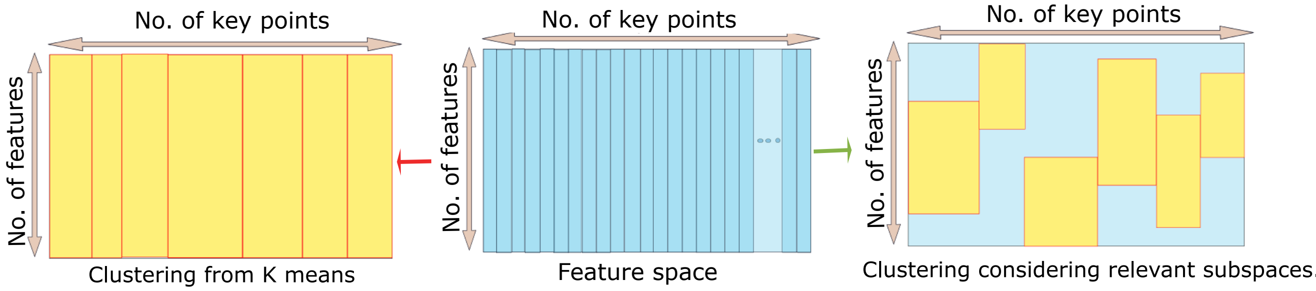

Figure 1. Motivation of the work: Figure in the middle describes the descriptor space with rows being the feature dimensions and the columns being the data points. The left figure delineates the output of using the K-means which considers all the feature dimensions. The figure on the right describes the output considering the relevant subspaces for each of the clusters which is our aim.

novel framework for using a supervised spectral subspace clustering method to build a more discriminative visual dictionary than the traditional one. The motivation of the work is also delineated in fig. 1.

SSC algorithm is based on the sparse representation of a data point i.e. each data point in a union of subspaces can be written as a sparse linear or affine combination of all other points. SMCE algorithm also utilizes the sparse representation of subspace data and model the optimization problem to find solution among close neighborhoods such that the points which are in close proximity get lower penalty. Unlike the previous two methods, SLBF algorithm uses Principal Component Analysis (PCA) to fit a subspace around each point considering some optimal neighborhood around the point. Then it constructs similarity matrix according to the distance between each pairwise subspaces.

There are several advantages of using SLBF method over other candidate methods for our purpose. For example, SLBF method is computationally less expensive as it does not use the expensive sparse optimization method for each point to find subspace members. Moreover, it is not necessary to define any influential parameter to tune as an input in SLBF while regularization parameter has to be tuned for methods using sparse optimization. In the following sections, we gradually discuss our proposed framework for forming visual dictionary, the experimental results and conclusion of the paper.

2. Proposed framework of using subspace clustering for dictionary creation

The method starts by taking the N feature descriptors, $\{x_i\}_{i=1}^N$ to build the visual dictionary where $x_i \in \mathbb{R}^D$. The flowchart of the proposed framework is shown in figure 2. It has two broad steps: subspace clustering and intrinsic subspace dimension estimation. A way of adding supervised information during the subspace clustering step is formulated and discussed later in section 2.2. However, estimation of the true subspace dimensions from the point cloud is a difficult problem on its own [9]. It is necessary to estimate the subspace dimensions for each cluster to quantize the image descriptors according to the subspace dimensions. Hence we propose a PCA based approach to get the relevant basis vectors for each subspace cluster.

2.1. Intrinsic subspace dimension estimation using PCA based approach

From subspace clustering, we get each of the total K subspace clusters, $\{S_i\}_{i=1}^K$ containing n_i number of points where $\sum_{i=1}^K n_i = N$. We have to get the intrinsic subspace basis vectors having dimension in $\mathbb{R}^{d_i}$ for each subspace.

At first, the eigenvectors for each subspace cluster is computed by using Singular Value Decomposition (SVD). The eigenvectors corresponding to the top few eigenvalues are selected as the basis of each subspace cluster. Thereafter, the Orthogonal Least Square(OLS) projective distance of the image descriptor to each subspace is calculated. If x is the new descriptor and P_{S_i} is the projection on subspace S_i, then the orthogonal least square distance (OLS) of x to S_i is calculated as (1).

$$\|x - P_{S_i}x\|_2^2 \tag{1}$$

Each descriptor is quantized to the subspace cluster for which OLS distance is the minimum over all the K clusters as shown in (2). Note that, the OLS distance is normalized by dividing with the number of eigenvectors, $\|P_{S_i}\|$ to discard the influence of selecting different number of eigenvectors for different subspaces.

$$\min_{S_i} \|x - P_{S_i}x\|_2^2 / \|P_{S_i}\| \tag{2}$$

Still, problem arises when for most of the clusters, the sample size, $n_i < D$, where D is the total number of feature dimensions. In that case, the computed eigenvectors are not always capable of reflecting the true underlying structure of the data due to having singular covariance matrix[2, 12]. Thus selecting more eigenvectors than n_i can create ambiguity in determining the true subspace[3, 12]. There are some studies and recommendations for selecting first few

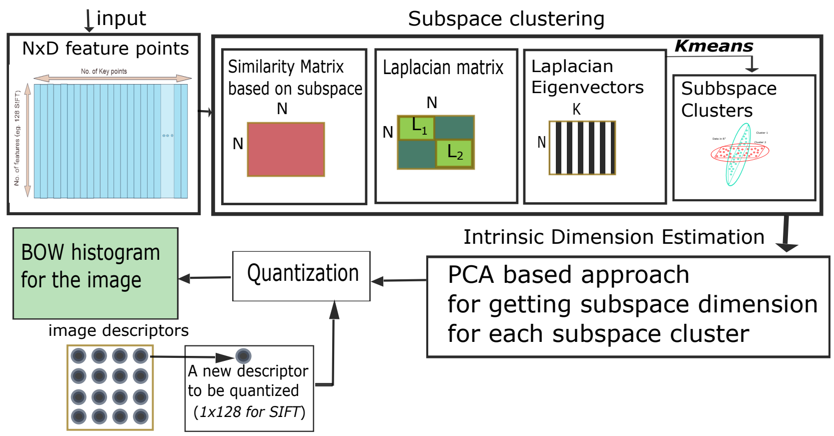

Figure 2. Flowchart of the proposed framework for building the visual dictionary using spectral subspace clustering.

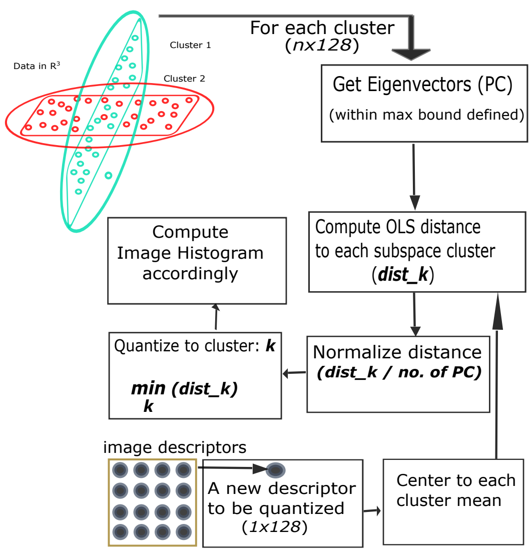

Figure 3. Building image Histogram from subspace clusters using PCA based approach.

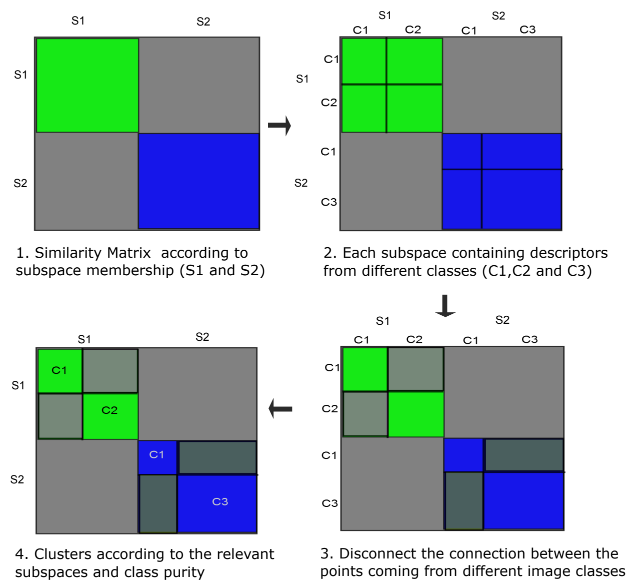

Figure 4. Steps of adding supervised information during spectral subspace clustering

eigenvectors which can show the subspace consistency for the underlying cluster [12]. This is why, we need to define a maximum limit on selecting eigenvectors for each subspace cluster such that the number of selected eigenvectors will be within the defined limit. For an example, we define maximum limit as 5 and there are two clusters with data matrix 20×128 and 3×128 where 128 is the feature dimension, we select only the first 5 eigenvectors for the first case and retain all the 3 eigenvectors for the later case. The flowchart of the whole procedure is shown in fig. 3.

2.2. Supervised Spectral Subspace Clustering

Our purpose of adding supervision for dictionary creation is to get visual words coming from relevant subspaces and from the same categorical images to retain the class purity. The first step involves building the similarity matrix W as discussed earlier which can be considered as a block diagonal matrix in terms of subspace similarity (step 1 in fig. 4). Each subspace contains descriptors from different categorical images (step 2 in fig 4). We simply disconnect the edge between the points coming from different image classes and retain other edge weights the same as before. In this way, we ensure to have clusters having descriptors from the same subspace and from the same categorical images. To remove the connection, the respective weights are made very close to 0. We chose 0.005 so that the graph does not get too sparse. Finally, the similarity matrix will have connected components in terms of the same subspace and the same image class. Note that, there can be clusters having the same class information but being in different subspaces (step 4 in fig 4) which also implies to have different distinc-

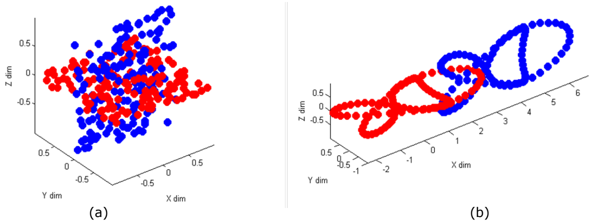

Figure 5. Synthetic Dataset: (a) with linear subspace and (b) with nonlinear subspace.

Subspace Clustering Method	Error (D1)	Error (D2)
K-means	49.67%	9.67%
SSC	10%	1.67%
SMCE	8.67%	0.3%
SLBF	0%	0.67%
SCC	0%	1%

Table 1. Missclassification error rate on synthetic datasets with linear (D1) and nonlinear (D2) subspaces.

tive features to identify the same categorical images.

3. Experimental Results

Two datasets have been used for the experiments – synthetic dataset with known ground truth and the 15 Scene Categories dataset[14] for image classification. The synthetic dataset has been generated with total 300 data points lying in $\mathbb{R}^3$. There are two axis aligned subspaces with each of them having 150 points lying in $\mathbb{R}^2$. One of the dimensions is shared by both subspaces. The dataset has been created for both linear (D1) and nonlinear (D2) subspaces (see figure 5). The performance of the subspace clustering methods are determined by the rate of misclassified subspace points calculated as (3). The 15 Scene Categories dataset consists of total 4885 images taken from 15 different scene categories. We select 100 images per class for training and the rest for testing. In total $28,000$ descriptors were randomly sampled from all the classes to build the dictionary.

$$\text{Error} \;=\frac{\text{no. of misclassified subspace points}}{\text{no. of total points}}\times 100\% \quad (3)$$

SIFT descriptors are used [15] along with the dense sampling method to have image patches. For dense sampling, 6 pixel shift in grid spacing is used and the size of the patch is kept 16×16. We used intersection kernel [16] along with linear SVM classifier by using LIBLINEAR package [10]. We cluster the descriptors into 1000 visual words for the image classification on the 15 Scene Categories dataset. All the experiments are run 10 times by reshuffling the test and training set. Finally, the average classification accuracy is reported for the performance measure.

3.1. Result discussion on Synthetic dataset

The experimental results on both synthetic dataset D1 and D2 are shown in Table 1. From the results, we can observe that the K-means clearly breaks the subspace structures while data points lie in a union of subspaces. On the contrary, the missclassification rates are quite lower for the subspace clustering methods than the K-means. For detecting clusters in linear subspaces, SLBF and SCC algorithm show the best performance having 0% missclassification rate. In case of nonlinear dataset (D2), all of the subspace clustering methods show very good results having missclassification rate within 2%. Interestingly, SLBF method exhibits better consistency for detecting both linear and nonlinear subspaces. Despite of having better results in synthetic data experiments, it was infeasible to run SCC algorithm in our further experiments with real world dataset due to its very high computational time and memory consumption. Finally, three methods – SSC, SMCE and SLBF are selected for further experiments on the real world dataset.

3.2. Result discussion on 15 scene categories

The result of applying spectral subspace clustering with PCA approach is shown in Table 2. We report maximum limits – 3, 5 and n_i on selecting the number of eigenvectors to show the varying influence on the results. For a solid comparison, The results of applying K-means along with PCA approach are also reported . It can be seen that SLBF method is dominating the result when first few eigenvectors (3 or 5) are selected. Note that, if we select number of eigenvectors equal to the number of data points per cluster (n_i), the performance becomes very poor due to the bad approximation of the subspaces. In case of using supervised approach (see results in the Table 3), all of the three methods outperform the baseline classification accuracy by selecting maximum 5 eigenvectors empirically.

4. Conclusions and Future Work

We propose a novel framework for using spectral subspace clustering for visual dictionary creation. From the experimental results, it can be observed that the SLBF algorithm works consistently better with our proposed framework and outperforms the baseline classification accuracy of BOW model. However, PCA approach for estimating the subspace basis dimensions has the limitation of tuning the maximum limit on selecting eigenvectors empirically. As future work, we plan to investigate a good solution to estimate the original subspace dimensions during subspace clustering in a coupled manner. Another experiment can be performed on using bi-level approach by at first employing subspace clustering to get clusters with good density and

No. of clusters	Max selected eigenvectors	Results (Spectral Subspace Clustering with PCA)			Kmeans with PCA	Baseline using K-means
		SMCE	SSC	SLBF		
1000	3	71.05	70.38	**73.76**	72.5	72.35
	5	71.52	69.47	**73.90**	72.06	
	n_i	59.00	16.18	32.29	51.97	

Table 2. Results on 15 scene dataset: Spectral Subspace Clustering with PCA approach (in %).

No. of clusters	Max selected eigenvectors	Results (Supervised Spectral Subspace Clustering with PCA)			Kmeans with PCA	Baseline using K-means
		SMCE	SSC	SLBF		
1000	3	72.80	72.76	**73.34**	72.5	72.35
	5	**73.13**	**73.06**	**73.91**	72.06	
	n_i	56.04	70.81	45.41	51.97	

Table 3. Results on 15 scene dataset: Supervised Spectral Subspace Clustering with PCA approach (in %).

then using the K-means per cluster with the assumption of having multiple clusters within each subspace.

References

[1] A. E. Abdel-Hakim and A. A. Farag. Csift: A sift descriptor with color invariant characteristics. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1978–1983. IEEE, 2006. 1

[2] J. Ahn, J. Marron, K. M. Muller, and Y.-Y. Chi. The high-dimension, low-sample-size geometric representation holds under mild conditions. *Biometrika*, 94(3):760–766, 2007. 2

[3] D. L. Bandalos and M. R. Boehm-Kaufman. Four common misconceptions in exploratory factor analysis. *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*, pages 61–87, 2009. 2

[4] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft. When is nearest neighbor meaningful? In *Database TheoryICDT99*, pages 217–235. Springer, 1999. 1

[5] S. Chandra, S. Kumar, and C. Jawahar. Learning multiple non-linear sub-spaces using k-rbms. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2778–2785. IEEE, 2013. 1

[6] G. Chen and G. Lerman. Spectral curvature clustering. *International Journal of Computer Vision*, 81(3):317–330, 2009. 1

[7] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, pages 1–2, 2004. 1

[8] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2790–2797. IEEE, 2009. 1

[9] E. Elhamifar and R. Vidal. Sparse manifold clustering and embedding. In *NIPS*, pages 55–63, 2011. 1, 2

[10] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. Liblinear: A library for large linear classification. *J. Mach. Learn. Res.*, 9:1871–1874, June 2008. 4

[11] G. Hinton. A practical guide to training restricted boltzmann machines. *Momentum*, 9(1), 2010. 1

[12] S. Jung, J. Marron, et al. Pca consistency in high dimension, low sample size context. *The Annals of Statistics*, 37(6B):4104–4130, 2009. 2, 3

[13] H.-P. Kriegel, P. Kröger, and A. Zimek. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(1):1, 2009. 1

[14] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2169–2178. IEEE, 2006. 1, 4

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 4

[16] S. Maji, A. C. Berg, and J. Malik. Classification using intersection kernel support vector machines is efficient. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2008. 4

[17] T. Tuytelaars, C. H. Lampert, M. B. Blaschko, and W. Buntine. Unsupervised object discovery: A comparison. *International journal of computer vision*, 88(2):284–302, 2010. 1

[18] R. Vidal. Subspace clustering. *Signal Processing Magazine, IEEE*, 28(2):52–68, 2011. 1

[19] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007. 1

[20] T. Zhang, A. Szlam, Y. Wang, and G. Lerman. Hybrid linear modeling via local best-fit flats. *International journal of computer vision*, 100(3):217–240, 2012. 1